

Does the explanatory gap rest on a fallacy?

Abstract

Many philosophers have tried to defend physicalism concerning phenomenal consciousness, by explaining dualist intuitions within a purely physicalist framework. One of the most common strategies to do so consists in interpreting the alleged “explanatory gap” between phenomenal states and physical states as resulting from a *fallacy*, or a *cognitive illusion*.

In this paper, I argue that the explanatory gap does not rest on a fallacy or a cognitive illusion. This does not imply the falsity of physicalism, but it has consequences on the kind of physicalism we should embrace.

Introduction

It is widely recognized that phenomenal consciousness seems to pose a serious metaphysical problem to physicalism. Even though we may have numerous reasons to think that phenomenal states are nothing over and above physical states (such as brain states), this thesis continues to strike us as extremely counter-intuitive. For example, I may be convinced by various arguments that experiencing pain is nothing but being in a certain brain state (say, C-fiber activation). However, when I focus introspectively on my current headache, I find myself deeply puzzled by this identity. How can this experience, this subjective feeling, be exactly the same thing as the activation of my C-fibers? When we experience this deep puzzlement, it is said that we face the *explanatory gap* (Levine, 1983, 2001), or that we have a strong *intuition of distinctness* (Papineau, 2002) regarding phenomenal and physical states.

Many dualist philosophers have transformed this intuition into *arguments*, designed to show that phenomenal states are indeed distinct from physical states. Physicalist philosophers, on the other hand, have employed different strategies to deal with this intuition. One of the most widespread strategies amounts to understanding this explanatory gap as the result of a *fallacy*, or a *cognitive illusion*. When it seems to us that phenomenal states cannot be identical with physical states (though they are), we are the victim of a fallacy – we succumb to a kind

of reasoning mistake. Various theories have been suggested to explain why we systematically commit such a fallacy (Loar, 1997; Papineau, 1993, 2002, 2007; Tye, 1999).

My goal is to argue against this kind of account. My strategy will be to carefully describe some examples of psychological processes underlying *fallacies* on the one hand, and *valid reasoning* on the other hand. I will point out some psychological features that are distinctive of these two kinds of processes. I will then turn to the process underlying the intuition of distinctness, and I will show that it much more closely resembles the process usually underlying a case of valid reasoning than the process which is typical of a fallacy. This gives us a reason to think that the explanatory gap is not the result of a fallacy. However, even if we accept this conclusion, it does not mean that we have to grant that we really enter phenomenal states which are metaphysically distinct from physical states – or that we have to deny that we face an explanatory gap. Indeed, there are alternative accounts of the intuition of distinctness, which do not interpret this intuition as the result of a fallacy, and yet preserve physicalism regarding the human mind.

In the first section, I will describe “Fallacy Accounts”, *i.e.* physicalist views that see the intuition of distinctness as the result of a fallacy. David Papineau’s account will be described in detail, as it is perhaps the most typical and the most elaborate example of such an account. In the second section, I will show how a physicalist can recognize the existence of the intuition of distinctness without necessarily seeing it as the result of a *fallacy*. I will describe two physicalist alternatives to Fallacy Accounts: “Introspective Illusion Accounts” and “Lack of Understanding Accounts”. In a third section, I will point out some typical psychological features of the processes underlying *fallacies* or *cognitive illusions* (as opposed to valid reasoning), using two examples. In the fourth section, I will show, on the basis of the previous analysis, that the process leading to the intuition of distinctness looks much more like a valid reasoning than a fallacy. I will then (section five) examine an objection to my argument, to which I will give a response, and I will finish off with some concluding remarks.

1. The explanatory gap as the result of a fallacy

Phenomenal states are states such that *there is something it is like* to be in them. A visual sensation of green, a gustatory sensation of chocolate, a burning sensation of pain on the forearm, etc., are typical examples of phenomenal states. These states bear *phenomenal properties*: properties of these states in virtue of which there is something it is like to be in them. For example, a visual sensation of green is a phenomenal state. It has a phenomenal

property, that we can call “phenomenal greenness”, in virtue of which it is a phenomenal experience of green.¹

Many philosophers of mind are physicalists: they think that all the mental properties we instantiate are nothing but physical properties – “physical properties” being here (as well as in the remainder of this paper) taken in an extended sense, so that it includes a vast set of properties (physical properties strictly speaking, but also physically realized functional properties, physically grounded properties, properties which logically supervene on physical properties, etc.). Physicalism, understood in this liberal sense, is a majority position within the field of philosophy of mind.²

However, even convinced physicalists are often puzzled by such a thesis in the case of phenomenal properties and phenomenal states. For example, let’s take the instantiation of phenomenal greenness by my current visual experience, as I look at a green apple. According to physicalism, this instantiation is supposed to be nothing over and above the instantiation of a certain physical property P. P can be a low-level biological property or a high-level physically realized functional property (or yet another kind of physical property), depending on the exact kind of physicalist theory one considers. However, it seems that, in all these cases, we can keep asking the question: how come the instantiation of P *feels like this* – how come it is the same thing as phenomenal greenness (rather than, say, phenomenal redness)? And it seems that we have no satisfying answer to this question, so that something about phenomenal greenness is left unexplained when we identify it with property P: there is an explanatory gap between the physical (broadly construed) and the phenomenal (Levine, 1983, 2001).

Philosophers understand this explanatory gap in various ways. Some influential thinkers, such as David Chalmers and Frank Jackson (Chalmers, 2009; Chalmers & Jackson, 2001) have argued that the “explanatory gap” has to be understood as a lack of *a priori* derivation from physical truths to phenomenal truths. According to them, there is an explanatory gap between the physical and the phenomenal because we cannot deduce the phenomenal truths about the world from the physical truths about the world. However, many philosophers disagree with such a diagnosis: they argue that the lack of *a priori* derivation may be *necessary* for the existence of the explanatory gap, but that it is not sufficient. For

¹ Here I take phenomenal properties to be properties of *mental states*. Some philosophers prefer to think about them as properties of *subjects* (experiencing subjects). I do not think anything substantial for my paper bears on this distinction; what I say in this paper could be restated in this alternative framework.

² The reasons to accept physicalism have generally mostly to do with causal considerations (Levine, 2001, Chapter 1; Papineau, 2002, Chapter 1). I won’t expound them here, as my goal is not to argue in favor of physicalism.

example, it can be argued that many non-phenomenal truths (such as truths involving proper names or indexicals) cannot be *a priori* derived from physical truths, and yet the corresponding identity statements (such as “Saul Kripke is the greatest living philosopher” or “Today is Tuesday”) do not give rise to an explanatory gap similar to the one we face in the case of consciousness (Papineau, 2008, p. 59). Some philosophers insist that even ordinary truths about water, for example, cannot be deduced from truths about H₂O (Block & Stalnaker, 1999) – and yet we usually accept the scientific identity statement “water is H₂O” without feeling that something substantive about water is “left out” by such an identity statement – contrary to what happens in the case of consciousness (Levine, 2001; Papineau, 2002, 2008, 2011). This line of reasoning has been very much debated. However, those who accept it can conclude that the distinctive explanatory gap we encounter in the case of consciousness must consist in something more than a mere absence of *a priori* derivation from the physical to the phenomenal. According to Papineau, the explanatory gap in fact really consists in what he calls the more basic “intuition of distinctness”: when we consider the instantiation of phenomenal greenness and the instantiation of property P, they simply *seem to be different things* (Papineau, 2002, 2011). Levine makes a similar diagnosis (Levine, 2001, Chapter 3), and thinks that the explanatory gap comes from the fact that we have a certain grasp of the nature of, say, phenomenal greenness, and that this property seems to us to have a nature that none of the physical properties we can think of has. He writes: “Whether we think of [the explanatory gap] as an explanatory gap or a distinctness gap, the problem is really the same” (Levine, 2007, p. 148).

In what follows, I will endorse a view of this kind: I will follow Levine and Papineau and I will presuppose: (1) That we do face an explanatory gap whenever we consider the claim that the instantiation of a given phenomenal property is nothing over and above the instantiation of a given physical property (broadly construed)³, and (2) that this explanatory

³ An anonymous reviewer pointed out the fact that the explanatory gap/intuition of distinctness does not arise with the same force (if it arises at all) with *some* physicalist theories of consciousness, such as, for example, Tye’s representationalist theory (Tye, 1995) which states that the instantiation of phenomenal greenness by one of my mental states is nothing but the instantiation (by one of my brain states) of the property of non-conceptually representing green, in a way that makes this representation available for my executive and thought system. I agree that the intuition of distinctness does not arise as strongly (if it arises at all) when we identify phenomenal properties with representational properties (compared to what happen when we identify phenomenal properties with low-level physiological properties of brain states, for example), but I do think that the intuition of distinctness arises in all its force when we specify that the concerned representational properties are *physical properties*. For example, I do think that the intuition of distinctness arises in all its force when we try to identify the instantiation of phenomenal greenness with the instantiation of a complex physical property which causally covariates in the appropriate way (or has causally covariated in the appropriate way through the evolutionary history of my ancestors) with such and such reflectance property of surfaces, and plays the appropriate functional role in my cognitive system (to take standard examples of the way in which physicalists conceive of

gap is the same thing as the intuition of distinctness, that it consists in the fact that the instantiation of a phenomenal property seems to be *different* from the instantiation of a physical property (broadly construed) and that (*pace* Chalmers and Jackson) it is not reducible to some other epistemic feature, such as a lack of *a priori* derivation from the physical to the phenomenal. And, from now on, I will use the terms “explanatory gap” and “intuition of distinctness” interchangeably.

Some philosophers have tried to defend physicalism against the pull of this intuition, by showing that we should expect such an intuition of distinctness between physical states and phenomenal states to arise even if physicalism is true. They explain this intuition by appealing to certain purely physical features of some of the concepts we use to think about phenomenal states: the so-called “phenomenal concepts”, that are notably, but not only, applied through introspection. This way of defending physicalism has been labeled the “Phenomenal Concept Strategy” (Stoljar, 2005), and it has been recently developed in many versions (Aydede & Güzeldere, 2005; Balog, 2012; Diaz-León, 2008, 2010, 2014; Elpidorou, 2013, 2016; Hill, 1997; Hill & McLaughlin, 1999; Levin, 2007; Loar, 1997; Papineau, 1993, 2002, 2007; Schroer, 2010; Sturgeon, 1994, 2000; Tye, 1999).

Many theorists amongst those following this strategy have suggested that the intuition of distinctness should be interpreted as the result of a “fallacy” (Papineau, 1993, 2002, 2007), or a “cognitive illusion” (Tye, 1999). In their view, our phenomenal concepts have some peculiar features. These features are such that we are systematically led to commit a *mistake* when reflecting on the relationship between phenomenal states and brain states: we are mistakenly led to judge that phenomenal states are distinct from brain states, even though they are not.

David Papineau, for example, states that phenomenal concepts present a “use/mention feature”: every occurrence of a given phenomenal concept involves the instantiation of the phenomenal property this concept refers to, or at least of a similar property. Therefore, every time I think about a certain type of phenomenal state, using phenomenal concepts, I activate a version of this experience (or at least a “faint copy” (Papineau, 2002, p. 118) of this experience).⁴ What happens then when we consider whether a certain phenomenal state (say, an experience of pain) is identical with a certain physical state (say, a C-Fiber activation)? We

representational properties). So, in my view, identifications of phenomenal properties with representational properties only escape the explanatory gap (if they do) when these representational properties are not explicitly conceived of as physical properties. I want to thank the anonymous reviewer for pointing out this difficulty.

⁴ Papineau has changed the details of his theory over the years, though he maintained his general line of thought (Papineau, 1993, 2002, 2007).

make use of two very different concepts: one of them is a phenomenal concept, whose application activates an experience of pain (or a copy of it). The other is a descriptive (physico-biological) concept which does not bring about any experience of this kind when it is applied – and the same would allegedly be true of any physical concept (broadly construed). Therefore, one way of thinking about pain has a distinctive feeling: *it is like something* to think about pain with a phenomenal concept. On the other hand, when I think about pain as C-fiber activation, there is no distinctive feeling associated with my thought. For this reason, as Papineau says, “there is an intuitive sense in which exercises of material concepts ‘leave out’ the experience at issue. They ‘leave out’ [...] the technicolour phenomenology, in the sense that they don’t activate or involve these experiences” (Papineau, 2002, p. 170).

This is where we commit the fallacy that Papineau calls the “Antipathetic Fallacy”:⁵ we systematically tend to “project” this phenomenological difference between our two thoughts on the *referents* of these two thoughts. In other words, we cannot help inferring, from the fact that our physical understanding of pain “leaves out” something when compared with our phenomenal understanding of the same thing, that the first must refer to something *different* from the second – *that what is left out is the referent itself*, or at least some features of the referent. This is why it appears to us that phenomenal concepts and concepts of brain states *must* refer to different states; that phenomenal states and brain states *must be distinct*. This explains the arising of the intuition of distinctness.

Michael Tye gave a somewhat (though not exactly) similar explanation of the explanatory gap (Tye, 1999, p. 712-713), even if he did not call the process by which the dualist intuition arises a “fallacy”, but a “cognitive illusion”. Here I consider that “fallacy” and “cognitive illusion” refer to the same kind of psychological process: a *mistaken* process, where the mistake takes place at the level of *reasoning*. Fallacies and cognitive illusions can notably be differentiated from *perceptual-like* illusions: the first take place at the level of reasoning (the manipulation of conceptual representations), and concerns the way in which we infer (wrongly) something from something else. The second take place at a lower-level and does not involve a mistake made when manipulating conceptual representations: if they rely

⁵ By using this term, Papineau makes a reference to the “Pathetic Fallacy”, described by the critic John Ruskin – the fallacy by which we tend to falsely attribute mental states that are our own to inanimate objects. When we commit the “Antipathetic Fallacy” described by Papineau, on the other hand, we (falsely) reject to attribute phenomenological properties to purely physical states.

on *inferences*, these inferences are necessarily only *subpersonal*, and the “premises” are encoded in a non-conceptual format, and have a non-conceptual content.

Amongst the proponents of the Phenomenal Concept Strategy, Brian Loar and Katalin Balog also made a similar proposal, and both described the explanatory gap as an “illusion” (Balog, 2012; Loar, 1997, p. 30-31). Although they do not specify the kind of illusion it is, I think that they probably have in mind a kind of *cognitive* (rather than perceptual, or perceptual-like) illusion.⁶ All in all, I think it is fair to say that explaining the explanatory gap as the result of a “fallacy” or a “cognitive illusion” is quite mainstream in recent philosophy of mind.

For reasons of simplicity, I will now call views that see the explanatory gap as a result of a fallacy “Fallacy Accounts”. Papineau’s theory is, to my mind, the most typical example of a Fallacy Account.

2. If not a fallacy, then what?

The intuition of distinctness is the intuition that phenomenal states are metaphysically distinct from physical states. If this intuition is correct, and if we really enter phenomenal states, it means that physicalism is false regarding the human mind. Fallacy Accounts avoid this conclusion by stating that the intuition is *not* correct, and that it is the conclusion of a fallacy. In such a view, physicalism is true; we really enter in phenomenal states (such as presented to us through introspection), and our understanding of physical states, on the other hand, is broadly correct. However, the peculiarity of phenomenal concepts makes it so that, when we *reflect* on the nature of phenomenal states and on their relations to physical states, we make a systematic cognitive mistake: we systematically judge wrongly that phenomenal states are *not* physical. And it is here, at the cognitive level – at the level of *reasoning*, the manipulation of conceptual representations – that the mistake arises. This mistake belongs to the category of *fallacies* or *cognitive illusions* which are studied by psychologists of reasoning (Kahneman, 2012; Pohl, 2004; Peter Wason & Johnson-Laird, 1972).

⁶ Loar describes this illusion as something that arises during “our philosophical ruminations”, which seems to confirm that what he has in mind is a *cognitive illusion*, rather than a perceptual one. It is perhaps less clear in Balog’s writing, as she sometimes seems to understand the illusion of the explanatory gap as something similar to a perceptual illusion. However, she does not take a clear stance on that question, and she does not explicitly analyze the explanatory gap as resulting of a kind of *perceptual-like* illusion (with everything that comparison implies).

However, there are at least two ways to preserve physicalism while granting that (a) we do face an explanatory gap but (b) this explanatory gap is *not* the result of a fallacy. The first corresponds to what I call “Introspective Illusion Accounts”. These accounts tell a story along the following line: we introspectively represent phenomenal states, and on the basis of this introspective process we apply phenomenal concepts. However, introspection is systematically inaccurate: it *misrepresents* phenomenal states. For example, it may represent phenomenal properties as having a qualitative nature that they don’t have, and that nothing physical (and therefore nothing real) has (Pereboom, 2009, 2011). The judgments we then form, on the basis of introspection, when we apply phenomenal concepts, may be seen as *systematically* mistaken: they retain the illusory component that arose at the introspective level. When we start to *reason* about the referents of phenomenal concepts, we conclude that phenomenal states, understood as the states which are exactly as presented to us through introspection, cannot be identical with physical states. But this reasoning is not mistaken, and it is not a fallacy. In fact, it is quite the opposite: this reasoning is perfectly correct. In accounts of this kind, it is *true* that phenomenal states (as presented to us through introspection⁷) are metaphysically distinct from physical states. However, it just happens to be the case that *we never are in any phenomenal states* (at least, not if we understand, by “phenomenal states”, states which are exactly as presented to us through introspection). In accounts of this kind, the intuition of distinctness is not the result of a fallacy, and in a way it is perfectly correct: phenomenal states “really” are distinct from physical states. But, given that we never enter phenomenal states, this does not endanger physicalism about the human mind. I take Pereboom’s view (a view he dubbed the “qualitative inaccuracy hypothesis”) to be a typical example of an Introspective Illusion Account, even though there are other (mostly scientific) theories that could fit this category (Graziano, 2013; Humphrey, 2011).

Physicalists who want to grant that we face an explanatory gap but deny that this explanatory gap results from a reasoning mistake can take another route, which corresponds to what we can call “Lack of Understanding Accounts” (for the term, see P. Sundström, 2008, 2011).⁸ Contrary to Introspective Illusion Accounts, Lack of Understanding Accounts endorse *realism* as well as physicalism about phenomenal states; they also state (like Fallacy

⁷ This caveat is important, because theories which endorse this explanation could try to say that phenomenal concepts have a *dual content*, following Chalmers’ distinction between edenic and ordinary content (Chalmers, 2006). That’s exactly what Derk Pereboom suggests (Pereboom, 2011). On this view, only phenomenal states understood as states satisfying phenomenal concepts’s edenic content would be distinct from physical states – while in some other understanding phenomenal states (understood as states satisfying merely phenomenal concepts’s ordinary content) could very well be identical with physical states.

⁸ I want to thank an anonymous for attracting my attention to this kind of view.

Accounts) that the intuition of distinctness is *incorrect*. However, this incorrect intuition is not seen as the result of a fallacy. The root of the intuition is not to be found in the fact that we make a *mistake* when we manipulate conceptual representations of physical and phenomenal states, but in the fact that we *lack* the conceptual representations that are required for understanding the physical nature of consciousness. Our introspective judgments are (generally speaking) true, and phenomenal states and properties really are nothing over and above physical states and properties. However, we do not have at our disposal the concepts which could allow us to make this fact intelligible. For this reason, we are led to judge that phenomenal states simply cannot be the same as physical states.⁹ Lack of Understanding Accounts have been defended in different forms by various authors in the last decades (McGinn, 1989; Nagel, 1974, 2000, Stoljar, 2005, 2006).

I described two physicalist alternatives to Fallacy Accounts: Introspective Illusion Accounts and Lack of Understanding Accounts. In both kinds of accounts, physicalism is true, and we *do* face an explanatory gap; but this explanatory gap is not the result of a fallacy. I do not claim that these two accounts exhaust the logical possibilities here: there might be other potential physicalist views which state that we face the intuition of distinctness while denying that it is the result of a fallacy.¹⁰ However, these two accounts seem to be the most developed alternatives to Fallacy Accounts. This means that, by arguing against Fallacy Accounts, we may give some indirect weight to the disjunction of these two views. This is to be noted, particularly for those who are primarily interested in the metaphysics of consciousness. Indeed, wondering about the correct account of the source of the explanatory gap can be seen as an interesting question in itself; a question that deserves to be answered for its own sake. However, it is important to note that the choice between the accounts previously

⁹ According to this kind of account, the primary root of the intuition of distinctness lies in the fact that we lack the adequate concepts to think about the relation between consciousness and the physical, and not in the fact that we commit a fallacy when reflecting on such a relation. However, there are interpretations of such accounts that combine both explanations. For example, it could be that, because we lack the adequate concepts to think about the relation between consciousness and the physical, we tend to manipulate the concepts we *do* have in a way that is inadequate and mistaken, which in turn gives rise to the intuition of distinctness – so that in some sense a kind of *reasoning mistake* plays an important role in the birth of the intuition. On such interpretations of Lack of Understanding Accounts, they are harder to distinguish from Fallacy Accounts – even though a distinction can still be drawn between views that think of fallacies as the ultimate source of the intuition (Fallacy Accounts properly speaking), and views that see them as just one element amongst others in the chain that leads to the intuition, the first element of the chain being our lack of the adequate concepts (Lack of Understanding Accounts so interpreted). Given that Lack of Understanding Accounts are not my primary topic here, I will not go further into the discussion of this kind of interpretations. Thanks to an anonymous reviewer for attracting my attention to this point.

¹⁰ And, of course, some physicalists simply deny that we face an explanatory gap in the sense in which I used the term. For example, some deny its existence (or at least its persistence on reflection) while others think that the explanatory gap does not exist *as I understand it*, given that it consists in nothing but the absence of a *a priori* derivation of phenomenal truths from physical truths.

described might also have consequences regarding the metaphysical nature of consciousness. The three accounts I describe are indeed all physicalist, but each suggest a different kind of physicalism. If Fallacy Accounts are correct, introspection is reliable: we really are in the kind of mental states that introspection presents us (even though we tend to commit mistakes concerning their metaphysical nature). That naturally leads to a *reductionist* view of consciousness, in which phenomenal consciousness is *real* and can be equated with a physical process of a kind we can currently understand (even if we may ignore its precise nature). But if Introspective Illusion Accounts are correct, introspection is inaccurate, and we never are in mental states such as the mental states that introspection presents to us. We are then led to an *illusionist* conception of consciousness, in which we deny the existence of genuine phenomenal states (Frankish, 2016): phenomenal consciousness is the result of an introspective illusion, and we *never are* in states that are like what is presented to us through introspection.¹¹ Finally, if Lack of Understanding Accounts are correct, we really enter phenomenal states, and these phenomenal states are purely physical, but some aspects of physical reality (relevant for consciousness) simply escape our understanding. Physicalism is true, but there is something we do not grasp about the physical, which makes physicalism about consciousness partially unintelligible. We may then be led towards what has sometimes been called (mostly by critics) a *mysterianist* conception of consciousness (Flanagan, 1991).

3. Fallacies and valid reasoning

I will now argue against Fallacy Accounts. My strategy will be to analyze the psychological process that gives rise to the explanatory gap, and to show that this process displays more similarities with processes underlying cases of *valid reasoning* than with the processes underlying *fallacies*. This will constitute an argument *against* the idea that the explanatory gap arises from a fallacy. However, as I have shown that there are physicalist alternatives to Fallacy Accounts, I do not think that my argument can be directly used against physicalism regarding the human mind. Besides, one important *caveat* to keep in mind: my own sympathy lies with Introspective Illusion Accounts, and given this fact it may very well be that the way in which my argument is set will seem to hint towards accounts of this kind. But I should be clear that I simply intend to argue *against* Fallacy Accounts here; I do not

¹¹ Illusionism does not strictly speaking imply eliminativism (at least in a certain sense of “eliminativism”), as an illusionist can always suppose that phenomenal concepts *do* refer, in the sense that their *ordinary* content is satisfied (even though their *edenic* content is not).

claim that my argument gives *more* indirect weight to Introspective Illusion Accounts than to Lack of Understanding Accounts (or, for that matter, than to other potential alternative accounts of the explanatory gap).

Let's start by examining a classic fallacy studied in psychology of reasoning, called the "conjunction fallacy" (Fisk, 2004; Kahneman, 2012; Tversky & Kahneman, 1983). A character is described to subjects in the following way: "Linda is thirty-one years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in antinuclear demonstrations". After that, subjects are asked to say which of the two following is the most probable: (1) Linda is a bank teller and (2) Linda is a bank teller and is active in the feminist movement.

It is logically impossible for (2) to be *more* probable than (1), and it is necessarily *less* probable (if we suppose that there is even a very small probability that Linda is a bank teller *not* active in the feminist movement, which is obviously true). There is therefore no doubt that (1) is the right answer. However, according to various repeated studies, about 85 to 90% of people who took this test (undergraduate students at major universities) answered that (2) was more likely (Kahneman, 2012, p. 158). This is a classic example of a *fallacy*, or *cognitive illusion*.

My goal is not to give a general psychological account of fallacies in general, or a particular account of *this fallacy* – whether it is Kahneman's account, which relies on a *dual system* theory, or another account. I simply want to make a few remarks concerning the psychological process that people undergo when they fall prey to this fallacy. This description seems to me to be rather devoid of theoretical commitment, though I must say that these remarks match Kahneman's descriptions and would fit perfectly well in his account of cognitive illusions; they are also partly inspired by what Rüdiger Pohl says about defining features of cognitive illusions (Pohl, 2004, p. 2-3).

The first thing I want to point out is that our tendency to commit the conjunction fallacy is embodied in an *automatic* psychological process. This means that, even when we try to inhibit it, we cannot help having the tendency to judge that (2) is more probable than (1) – even if we can prevent from judging that (2) is more probable than (1). But it also means (and this is the point I am primarily interested in noting) that this fallacy is *not* the product of careful reflection. Careful reflection about the issue at stake does *not* give rise to the fallacious judgment. On the contrary, careful reflection is what allows us to find the right

answer, which is that (1) is more probable than (2) – even though reflection does not suppress the tendency to make the fallacious judgment. So: careful reflection certainly does not suppress the tendency to commit the fallacy (otherwise, this case would not be a genuine case of cognitive illusion, but simply a “normal” mistake), but it does not *produce* the fallacious result either.

The second thing I want to remark is that, even though there is obviously a strong and widely shared tendency to commit this fallacy, the *right answer* to the question is perfectly intelligible on careful reflection. By that, I mean that, on careful reflection, we are perfectly able to *identify* the right answer, and we are perfectly able to understand *why* the right answer is the right answer: why (1) is necessarily more probable than (2), given notably that (2) implies (1) but not conversely. We may still *tend* to judge that (2) is more probable than (1), but we will nevertheless find the right answer perfectly intelligible and unproblematic on careful reflection.

I take it that these two remarks are independent of any particular psychological account of fallacies (even if they fit well with Kahneman’s two systems theory), and that they would also apply to other typical fallacies and cognitive illusions (which I don’t describe here for reasons of space), such as – to take an example of a fallacy which is well-known by philosophers – the fallacy committed by a vast majority of subjects in the famous Wason Selection Task (Paul Wason, 1966).

I now want to describe a very different kind of psychological process. It is a case of *valid reasoning*, where the reasoning is made on the basis of false premises, and the false premises are obtained by way of an unreliable and illusory perceptual process. Let’s suppose that a woman, Anne, is about to enter a room. Some people tell her that there is nothing in the room but a white chair; let’s say she believes them. They happen to be telling the truth, as the room is really empty, except for a white chair. However, when Anne enters the room, she happens to have a hallucination of a black cat sitting on the white chair.¹² Let’s also suppose that, on the basis of this nonveridical visual experience, she forms a perceptual judgment of the kind: “there is a black cat in this room”.

At this point, a little bit of reasoning (which, in most real cases, will probably happen in an extremely quick way) can show her that her previous belief, according to which there is nothing in the room but a white chair, must be abandoned in the light of this new piece of

¹² Perhaps she has this hallucination because she just took (without knowing it) some very elaborate psychoactive drug which causes visual hallucinations, while leaving her reasoning capacity intact; or perhaps because she has been secretly equipped with a sophisticated TMS device that directly stimulates her visual cortex. It does not really matter here.

information. For example, she can reason like this: “There is a black object in the room, an object cannot be both black and white, so there is an object in the room which is not white, which means that it’s false that there is nothing in the room but a white chair”; or like this: “There is a cat in the room, something cannot be both a chair and a cat, so there is an object in the room which is not a chair, which means that it’s false that there is nothing in the room but a white chair”, etc. In other words: a little bit of reasoning leads her to consider a contradiction between her previous belief about the room (which happened to be true) and the content of her new judgment that there is a black cat, made on the basis of her fallacious visual experience. It may very well lead her to *abandon* her previous belief that there is nothing in the room but a white chair.

I now want to make a few remarks on the story I just described. The psychological process which leads Anne to have a visual experience of a black cat is misleading, and it gives rise to an illusory perceptual experience. On the basis of this illusory experience, she will then form a false belief concerning the presence of a black cat in the room. However, the illusory component here is not *cognitive*, but *perceptual-like*: it is her perceptual (visual) system which malfunctions here, and leads her to form this false belief. No reasoning mistake is involved here: she commits no *fallacy*. Notably, she makes absolutely no reasoning mistake when her (simple) reasoning allows her to understand that the belief that there is a black cat in the room contradicts the belief that there is nothing in the room but a white chair. This reasoning is perfectly valid: she is fully justified to infer, from the premise that there is a black cat in the room, that it is false that there is nothing but a white chair in the room – granted a few widely accepted definitional statements concerning colors, kinds of objects, etc. So, inasmuch as some reasoning is involved when she comes to abandon her previous true belief that there was nothing in the room but a white chair, this reasoning is perfectly valid – it does not consist in a *fallacy*, or in a *cognitive* illusion of any kind. However, the premise on which this reasoning is based is false, and it is obtained by way of a dysfunctional perceptual device. The illusory component here arises at the perceptual level: if there is an illusion here, it is a perceptual, not a cognitive illusion.¹³

¹³ Of course, there is another step between the perceptual illusion and the valid reasoning that leads to the conclusion that it is false that there is nothing in the room but a white chair: the step by which Anne “endorses” her (illusory) perception, and judges, on the basis of this perception, that there is a black cat in the room. However, I think that it is clear that this step cannot be seen as consisting in a kind of *reasoning mistake* – arguably, we commit no reasoning mistake when we endorse the content of our perception (as long as we do not have reasons to doubt our perceptions). Therefore, the existence of this step does not endanger the following point, which is crucial for my argument: in the case described, Anne commits no fallacy, and falls prey to no cognitive illusion.

I now want to point out some psychological features of the reasoning Anne uses to conclude, from the judgment that there is a black cat in the room, that it's false that there is nothing in the room but a white chair.

First, this reasoning is *sustained* by careful reflection. The more Anne (or anyone in her situation) reflects on this subject, the more it is obvious to her that, if there is a black cat in the room, then it is *impossible* that there is nothing in the room but a white chair.

Second, she can formulate many different specific arguments to reach the same conclusion. For example, as I showed previously, she can base her reasoning on the *color* of the objects considered (something cannot be both black and white, therefore if there is a black cat, then there is more than just a white chair), on the *kinds* of the objects concerned (a cat cannot be a chair), etc. Even though all of these arguments are extremely simple, it is important to note that we can easily come up with many of them.

Third, these arguments are deductive, which notably has the following consequence: even when Anne thinks hard (and, maybe, *especially* when she thinks hard), she simply cannot understand how the conclusion of these arguments could be false, granted that the premises are true. In this case: if Anne accepts that there is indeed a black cat in the room, and if she accepts some commonly shared definitional statements about the concepts she uses, then she cannot make sense of the idea that *it is still true that there is nothing in the room but a white chair*. No matter how hard she tries, she cannot apprehend or picture a situation that would make both of these beliefs true.

I just described two kinds of psychological processes: both lead at false conclusions. One is a case of *fallacy* (or *cognitive illusion*). The second is a case of valid reasoning, which leads to a false conclusion, and is based on a false premise (obtained through an illusory perceptual process). In the first case, the illusory component arises at the cognitive level. In the second case, it arises at a perceptual level. I tried to highlight some notable features which distinguish these two psychological processes. In what follows, I will focus on the psychological process which leads to the intuition of distinctness, and I will compare it to the two processes I just described, in order to argue against Fallacy Accounts.

4. The intuition of distinctness as the result of valid reasoning

I now want to focus on the psychological process which leads to the intuition of distinctness. My goal is to show that this process has more in common with the process underlying a typical case of valid reasoning than with the process underlying a typical fallacy.

I thus intend to support the thesis according to which the intuition of distinctness is not the result of fallacy.

Let's say that I focus on my current visual experience – for example, on my current visual experience of blue, as I am looking at the painting *Bleu II* by Joan Miró. I then try to think that this experience is *identical* with some electrochemical activity currently taking place in my visual cortex. At this point, I encounter an intuition of distinctness, I face the “explanatory gap”: it seems to me that the two things I am thinking about, and that I am trying to identify, *cannot really be identical*. I am deeply puzzled by their putative identity, and I am very reluctant to accept it, as it seems blatantly false.

Now, let's try to describe the psychological process by which I am led to judge, or at least to be strongly tempted to judge, that my phenomenal experience cannot be identical with a brain state. To begin with, one crucial thing is striking: careful reflection about the objects grasped by my two thoughts (my “phenomenal” thought and my “physical thought” – which according to the physicalist grasp the same object) *does support* the intuition of distinctness. The more I think introspectively about my current experience of blue, the more I meditate on its subjectivity, its qualitateness, the fact that it is directly felt and experienced, etc., the more it seems obvious to me that *it simply cannot be identical* with a blunt, objective, “blind” physical process (regardless of its complexity).

Of course, further reflection can convince me that physicalism is true. After all, if that were not the case, there would be no physicalists on Earth. However, this further reflection relies on other considerations: it is not merely based on my current grasp of the objects at hand. For example, this further reflection may rely on metaphysical considerations concerning ontological simplicity, or causality, etc. The important point is that, if I consider what I am led to believe simply by *carefully reflecting* on my current experience of blue on the one hand, and on an electrochemical cortical activity on the other hand, I find that such careful reflection *does lead me to the intuition of distinctness*: it leads me to the idea that my experience *cannot be identical* with a physical process (here, a brain process).

If indeed the intuition of distinctness is, as I claim, the product of careful reflection, then the psychological process that causes it is more similar to a case of valid reasoning than to a fallacy. Actually, I even think that the intuition of distinctness not only *is* the product of careful reflection, but also that *only* careful reflection produces it. By that, I mean that the intuition of distinctness does *not* arise at first glance, when we merely think quickly and superficially about the issue at hand. On the contrary, it only appears when we carefully

reflect on the nature of the entities we are thinking about.¹⁴ However, this point may be less easy to support, while I take it to be extremely plausible that careful reflection *does indeed* produce the intuition of distinctness. When I contemplate attentively my experience of blue, when I think about its various features with great care, I find myself to be very reluctant to identify it with a physical process.

I think that this important point has been often been missed by physicalists working on these questions, perhaps because they have tended to conflate two close issues which should be carefully distinguished. One issue is the nature of the process that leads to the *intuition* of distinctness (intuition according to which consciousness is not physical), and the other one is the nature of the process that leads to the *belief* that we enter phenomenal states which are not entirely physical, so that physicalism is false concerning the human mind. Of course, physicalists want to say that careful reflection leads us to *abandon* the belief that we enter phenomenal states which are not entirely physical, or at least that it *should* have this effect. After all, if they did not think so, why would they be physicalists? However, this should not preclude them from recognizing that the *intuition of distinctness* (which by the way is not a belief but rather a peculiar *disposition* to believe something which is triggered in certain conditions) *is indeed produced by careful reflection* on the concerned objects. We can accept this thesis, and yet think that, when we take into account *other considerations* (general metaphysical considerations, for example), we can be led by careful reflection to accept physicalism.

So, this first and crucial feature of the process by which we are led to the intuition of distinctness gives us a reason to think that this process is more akin to a case of valid reasoning, than to a fallacy. Let's now focus on two other features of this process.

When I focus on my current experience of blue, it seems to me that it cannot be the same thing as an electrochemical activity in my visual cortex. This is precisely the intuition of distinctness. However, if I carefully examine this situation, I find that I am not only *strongly disposed to believe* that the two entities are distinct. I also have a lot of difficulty

¹⁴ I have only anecdotal evidence supporting this claim: when I teach philosophy of mind to undergraduates, I find that, even if many of them are intuitive dualists, they are rarely dualists for reasons specifically related to the hypothetical irreducibility of phenomenal states. They are often reluctant to accept physicalism simply because it seems to them that, by treating human minds as "machines", physicalists cannot account for the creativity and the freedom that human beings possess. However, after some teaching and some thought experiments, which I think aim at triggering careful reflection on the objects considered, students often start to be puzzled by physicalism *concerning phenomenal consciousness in particular*, and they begin to encounter the intuition of distinctness as I understand it (even though they may very well accept physicalism for other reasons). I think that this anecdotal evidence weighs in favor of the claim that the intuition of distinctness is produced by, and *only* by, careful reflection on the concerned entities.

*understanding how they could be the same.*¹⁵ Even if I try really hard to *accept* this identity, I still am really puzzled and bewildered when I aim at representing *what it would be for this identity to be the case*. There is a sense in which the understanding of this identity systematically *eludes* me; a sense in which I simply don't find it fully intelligible.¹⁶ In other words: I am not only strongly "pulled" towards anti-physicalism, I am also having a hard time picturing what it would mean for physicalism to be true.

Of course, from the point of view of the physicalist, there *has* to be a way to think about physicalism that makes this doctrine perfectly intelligible, but this way may correspond to different conditions of thought – for example, it may imply the use of other concepts. What I want to point out is that, when I introspectively focus on one of my current experiences, it seems to me that this experience is distinct from a brain activity, and it is very difficult for me to fully apprehend how these two things *could be identical*. So, in this respect, too, the psychological process that leads to the intuition of distinctness resembles valid reasoning more than a fallacy, as we find ourselves in trouble when we try to simply *understand* how a certain situation can be true: the fact that my mind is entirely physical while I have subjective and qualitative experiences in one case, or the fact that there is nothing but a white chair in the room while there is a black cat in Anne's case.

Finally, there is a third point I want to highlight. When I focus on my current experience of blue, many reasoning paths can lead me to the intuition of distinctness, i.e. to the idea that this experience cannot be identical with a brain state. For example, I can focus on the *qualitative character* of my experience: nothing in my brain has such a qualitative character, so the two things cannot be identical (by Leibniz's Law). I can follow the same reasoning based on its *subjective character*, the fact that this experience of blue is inherently *for me*.¹⁷ I can also reason on the basis of epistemological or modal considerations, those operative in the Knowledge Argument (Jackson, 1982), or the various Modal Arguments (Chalmers, 1996;

¹⁵ It is important to note that these two psychological facts are quite distinct. Indeed, being strongly disposed to believe that P and having difficulties understanding how not-P could be true, are two different things and the former does not imply the latter. Consider for example the following fact: my current visual experience of my two hands very strongly disposes me to believe that I have two hands. However, I have no difficulty understanding how, in spite of what I experience, I may *not* have two hands (for example, I can picture a situation in which I am hallucinating); I have no problem apprehending this possible situation (Kammerer, 2018).

¹⁶ This is why many people find that the most tempting thing to say when facing physicalism is simply, as Joseph Levine wrote in conclusion of his review of Christopher Hill's (materialist) book on consciousness: "believe it if you can" (Levine, 2011).

¹⁷ For the distinction between qualitative character and subjective character, see (Kriegel, 2005; Levine, 2001, p. 7-9).

Kripke, 1980), to reach the same conclusion.¹⁸ In this respect, the psychological process by way of which I reach the intuition of distinctness resembles the case of valid reasoning previously described. In the example I gave (Anne's case), the subject could follow many paths to conclude, from the fact that there is a black cat in the room, that it cannot be the case that there is nothing in the room but a white chair. On the other hand, fallacies and cognitive illusions function quite differently. In a fallacy, such as the conjunction fallacy described by Kahneman and Tversky, we simply *jump* to the fallacious conclusion, in a single and simple step, which is difficult to analyze and which does not seem to allow for much variations (at least, not much carefully conducted and examined variations). So, this last feature of the psychological process underlying the intuition of distinctness gives us another reason to think that this intuition is not the result of a fallacy, but stems from something which is closer, from a psychological point of view, to a valid reasoning.

The psychological process which sustains the intuition of distinctness/explanatory gap is similar to the processes usually underlying valid reasoning, and dissimilar to the processes underlying typical fallacies. This gives us a reason to think that the intuition of distinctness is *not* the result of a fallacy (or a cognitive illusion). To the extent that this intuition involves some reasoning, we have reasons to think that this reasoning is valid. This gives us a reason to reject Fallacy Accounts

5. Objection and response, and conclusion

I have argued that the psychological process leading to the intuition of distinctness is closer to a case of valid reasoning than to a fallacy, and that this gives us a reason to reject the idea that this intuition is the result of a fallacy – an idea which is at the heart of Fallacy Accounts. Of course, this conclusion could be resisted. For example, it may be that the intuition of distinctness is caused by a fallacy of a very peculiar kind, endowed with some special psychological features in virtue of which it very much resembles a valid reasoning.

A proponent of Fallacy Accounts could notably point out that the views suggested by Papineau or Tye do not merely state that the intuition of distinctness is the result of a fallacy. They claim that it is the result of a peculiar fallacy, caused by the unique use/mention feature

¹⁸ Of course, physicalists reject the conclusion of such arguments, which mean that they have to say that *something* is wrong with these arguments. However, I think that the overwhelming majority of physicalists will grant that the problem with these arguments can hardly be understood as being simply a matter of fallacy, or cognitive illusion. Physicalists who reject these arguments have to reject one of the *premises* of these arguments (and they often go as far as to admit that these false premises still have some kind of *prima facie* rational plausibility).

of phenomenal concepts: any use of a phenomenal concept involves an activation of the experience it refers to. In such view, it is because the use of phenomenal concepts involves a peculiar phenomenology – while the use of physical concepts does not – that we are led to think (wrongly) that something is left out by physical concepts, and that the referent of a phenomenal concept cannot be nothing over and above physical states. Could it be that the peculiar way in which this fallacy is generated explains why it is psychologically more akin to a case of valid reasoning, than to a typical case of fallacy?

Perhaps one could argue like this: careful reflection on the identity of phenomenal states and physical states involves phenomenal concepts. For that reason, we should not expect the intuition of distinctness to *disappear* on careful reflection, as the cause of the intuition (phenomenal concepts and their peculiar features) is still present. So far, so good: this is what we should expect in normal cases of fallacies (which arguably might also be generated by some features of our representations, which subsist whether or not we are carefully reflecting). But then something peculiar might happen because of the unique use/mention feature of phenomenal concepts. Careful reflection might plausibly lead to an increase in introspective attention, and for that reason it could make the real phenomenological difference between phenomenal concepts and physical concepts *more salient and more obvious* for the subject. But the grasp of this phenomenological difference is supposed to be the cause of the fallacy leading to the intuition of distinctness. Its peculiar salience on careful reflection could then explain why careful reflection, not only does not eliminate, but also actively *sustains* this particular fallacy – making it superficially more similar to a case of valid reasoning than to a typical case of fallacy. Maybe this could also explain why even careful reflection does not make the right position (the position according to which experiences really just are physical states) appear as such, contrary to what happens in other fallacies.¹⁹

It is plausible that careful reflection (at least in certain conditions) leads to an increase in introspective attention, which could in turn highlight the real phenomenological difference between uses of phenomenal concepts and uses of physical concepts: when paying more closely attention to our own thinking, we might notice more clearly that thinking with a phenomenal concept involves a certain phenomenology, while thinking with a physical concepts does not. However, one should keep in mind that, according to Papineau and Tye, we do not commit a fallacy when we *accurately* notice that uses of phenomenal concepts and physical concepts have a different phenomenology. We only commit a fallacy when we

¹⁹ I want to thank the anonymous reviewer who inspired this objection.

project this difference in phenomenology *on the referents of the concepts*. But it is this very fallacious projection, I think, that we should expect *not to be sustained* by careful reflection (though careful reflection may intensify the real difference illegitimately projected). On the contrary, we should expect careful reflection (maybe precisely because it involves increased introspective attention) to make us able to distinguish what belongs to the way in which the use of the concepts *feels like*, and what belongs to what the concept represents about its referent. I do not see any compelling reason to think that the peculiar salience of the feeling associated with the use of one the concepts should prevent the making of such distinction – quite the contrary, as the introspective salience of this feeling should improve our ability to make this distinction. For this reason, the identity of phenomenal states and physical states should be made clear and intelligible during these moments of careful reflection, when we are able to tear apart what belongs to the use of the concepts and what belongs to their referents. However, it is not what happens: even when we engage in careful reflection, it seems to us that it is *experiences themselves* which have aspects (qualitativity, subjectivity, etc.) not had by physical states. Careful reflection sustains intuitions about *experiences* – not simply about how it feels to use certain concepts. Therefore, I do not think that proponents of Fallacy Accounts could appeal to the peculiar features of phenomenal concepts to explain why, even though the intuition of distinctness rests on a fallacy, it has much in common with a case of valid reasoning from a psychological point of view.

Of course, it is still possible for the proponent of Fallacy Accounts to maintain that the fallacy we commit in the case of the identity of phenomenal states and physical states is uniquely peculiar (for some *other* reason), so that this fallacy resembles a case of valid reasoning much more than it resembles other typical fallacies. However, such an account would have to be worked out and justified independently. I think that the burden of proof has been now shifted to the defender of Fallacy Accounts, if she wants to maintain that the intuition of distinctness really arises from a fallacy.²⁰

²⁰ The defender of Fallacy Accounts may suggest a view in which the fallacy does not arise at the level of the question “are phenomenal states identical with physical states?”, but at another (slightly different) level which concerns questions such as “can these features which introspection ascribes to phenomenal states, such as subjectivity, or “qualitative-ness”, be purely physical features?” In this view, we would commit indeed no fallacy when we judge that phenomenal states cannot be physical, from the premise that they are qualitative and subjective *and* the premise that subjectivity and qualitative-ness cannot be physical features. However, the fallacy emerged “earlier”, when we judged (fallaciously), when reflecting on subjectivity and qualitative-ness, that subjectivity and qualitative-ness cannot be physical features. However, I think that this view could be targeted by an argument extremely similar to the one I just gave in this paper. Indeed, it could be argued similarly that the kind of process by which we come to think that subjectivity or qualitative-ness cannot be purely physical features, for as much as it involves reasoning and the manipulation of conceptual representations, is much more similar to a process of valid reasoning than to a process of fallacy (cognitive illusion), pretty much

If rejecting to treat the intuition of distinctness as a fallacy meant that we were forced to infer that there is a real metaphysical gap between two different kinds of real processes, phenomenal processes and physical processes – that is, to infer that physicalism is false – then the cost of abandoning Fallacy Accounts would be very high. Physicalist philosophers would then certainly be tempted by all kinds of *ad hoc* moves in order to save Fallacy Accounts. However, as I tried to show previously, there are at least two serious alternatives for philosophers who want to account for the existence of a robust intuition of distinctness, while endorsing physicalism: Introspective Illusion Accounts and Lack of Understanding Accounts.

As I said, my own sympathy lies with Introspective Illusion Accounts. I think it may be visible in the way in which I described the typical example of valid reasoning leading to a false belief (Anne's example). Indeed, this example precisely mirrors what happens in the case of consciousness according to Introspective Illusion Accounts. Accounts of this kind state that we are mistakenly led to believe that physicalism is false, on the basis of a valid reasoning which requires a false premise obtained through an introspective (but not a cognitive) illusion; in the same way, Anne is mistakenly led to believe that it is false that there is nothing in the room but a white chair, on the basis of a valid reasoning which requires a false premise obtained through a perceptual (but not a cognitive) illusion. However, my goal here is simply to argue that the intuition of distinctness is not the result of a fallacy, and that Fallacy Accounts are false: I did not intend to argue for Introspective Illusion Accounts over Lack of Understanding Accounts (or other possible accounts).

Introspective Illusion Accounts and Lack of Understanding Accounts are not themselves without problems. Introspective Illusion Accounts seem to lead to illusionism about consciousness, which has been subjected to many objections (Balog, 2016; Frankish, 2016; Nida-Rümelin, 2016) and still receives a respectable amount of incredulous stares from philosophers. Lack of Understanding Accounts, on the other hand, are sometimes accused of constituting *unstable* versions of physicalism (Chalmers, 2002) – at times dangerously flirting with non-physicalist options, such as neutral monism. However, the fact that these two alternatives are available shows that one can accept that the explanatory gap is not the result of a fallacy while preserving one's commitment to physicalism. Given the problems

in the same respect as the process by which we come to think that phenomenal states cannot be identical with physical states (which I examined in detail in the paper). Indeed, (1) the process by which we come to judge that subjectivity and qualitative-ness cannot be physical *is indeed* the product of careful reflection; (2) it is very hard for us to understand how the content of the judgments at hand could be false (how subjectivity, for example, given the introspective grasp we have of it, could be a purely physical feature); (3) the same conclusion (that subjectivity or qualitative-ness cannot be physical) could be reached through several various arguments (we could reason on the categorical, modal, epistemological, properties of subjectivity or qualitative-ness, etc.). Thanks to Joseph Levine for raising this point in correspondence.

encountered by Fallacy Accounts and (as I hope to have shown) their lack of psychological plausibility, I think physicalist philosophers should devote more time and attention to these two conceptions, as well as to other potential alternatives.

References

- Aydede, M., & Güzeldere, G. (2005). Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness. *Noûs*, 39(2), 197-255.
- Balog, K. (2012). Acquaintance and the Mind-Body problem. In C. Hill & S. Gozzano (Ed.), *New Perspectives on Type Identity: The Mental and the Physical*. Cambridge University Press.
- Balog, K. (2016). Illusionism's Discontent. *Journal of Consciousness Studies*, 23(11-12), 40-51.
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, 108(1), 1-46.
- Chalmers, D. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
- Chalmers, D. (2002). Consciousness and its Place in Nature. In D. Chalmers (Ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford University Press.
- Chalmers, D. (2006). Perception and the Fall from Eden. In T. Szabo & J. Hawthorne (Ed.), *Perceptual Experience* (p. 49-125). Oxford: Oxford University Press.
- Chalmers, D. (2009). The Two-Dimensional Argument Against Materialism. In B. McLaughlin (Ed.), *Oxford Handbook of Philosophy of Mind*. Oxford University Press.
- Chalmers, D., & Jackson, F. (2001). Conceptual Analysis and Reductive Explanation. *Philosophical Review*, 110, 315-361.
- Dennett, D. (1988). Quining Qualia. In A. Marcel & E. Bisiach (Ed.), *Consciousness in Modern Science*. Oxford University Press.

- Dennett, D. (1991). *Consciousness Explained*. Penguin.
- Diaz-León, E. (2008). Defending the Phenomenal Concept Strategy. *Australasian Journal of Philosophy*, 86(4), 597-610.
- Diaz-León, E. (2010). Can Phenomenal Concepts Explain The Epistemic Gap? *Mind*, 119(476), 933-951.
- Diaz-León, E. (2014). Do a Posteriori Physicalists Get Our Phenomenal Concepts Wrong? *Ratio*, 27(1), 1-16.
- Elpidorou, A. (2013). Having it both ways: consciousness, unique not otherworldly. *Philosophia*, 41(4), 1181-1203.
- Elpidorou, A. (2016). A posteriori physicalism and introspection. *Pacific Philosophical Quarterly*, 97(4), 474-500.
- Fisk, J. (2004). Conjunction Fallacy. In R. Pohl (Ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgment and Memory* (p. 23-42). Hove, East Sussex: Psychology Press.
- Flanagan, O. (1991). *The Science of the Mind*. Cambridge, MA: MIT Press.
- Frankish, K. (2016). Illusionism as a Theory of Consciousness. *Journal of Consciousness Studies*, 23(11-12), 11-39.
- Graziano, M. (2013). *Consciousness and the Social Brain*. Oxford: Oxford University Press.
- Hill, C. (1997). Imaginability, Conceivability, Possibility and the Mind-Body Problem. *Philosophical Studies*, 87, 61-85.
- Hill, C., & McLaughlin, B. (1999). There Are Fewer Things in Reality Than Are Dreamt of in Chalmers' Philosophy. *Philosophy and Phenomenological Research*, 59(2), 445-454.
- Humphrey, N. (2011). *Soul Dust: The Magic of Consciousness*. Princeton: Princeton University Press.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32(April), 127-136.
- Kahneman, D. (2012). *Thinking, Fast and Slow*. Penguin.
- Kammerer, F. (2018). Can you believe it? Illusionism and the illusion meta-problem. *Philosophical Psychology*, 31(1), 44-67.

- Kriegel, U. (2005). Naturalizing Subjective Character. *Philosophy and Phenomenological Research*, 71, 23-56.
- Kripke, S. (1980). *Naming and Necessity*. Harvard University Press.
- Levin, J. (2007). What is a Phenomenal Concept? In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Levine, J. (1983). Materialism and qualia: the explanatory gap. *Pacific Philosophical Quarterly*, 64(October), 354-61.
- Levine, J. (2001). *Purple Haze: The Puzzle of Consciousness*. Oxford University Press.
- Levine, J. (2007). Phenomenal Concepts and the Materialist Constraint. In T. Alter & S. Walter (Ed.), *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Levine, J. (2011). Review de Consciousness, by Christopher S. Hill. *Mind*, 120(478), 527-530.
- Loar, B. (1997). Phenomenal States (Revised Version). In N. Block, O. Flanagan, & G. Güzeldere (Ed.), *The Nature of Consciousness* (p. 597-616). MIT Press.
- McGinn, C. (1989). Can we solve the mind-body problem? *Mind*, 98(July), 349-366.
- Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, 83(October), 435-450.
- Nagel, T. (2000). The Psychophysical Nexus. In P. Boghossian & C. Peacocke (Ed.), *New Essays on the A Priori* (p. 432-471). Oxford: Clarendon Press.
- Nida-Rümelin, M. (2016). The Illusion of Illusionism. *Journal of Consciousness Studies*, 23(11-12), 160-171.
- Papineau, D. (1993). Physicalism, Consciousness, and the Antipathetic Fallacy. *Australasian Journal of Philosophy*, 71, 169-183.
- Papineau, D. (2002). *Thinking about Consciousness*. Oxford University Press.

- Papineau, D. (2007). Phenomenal and Perceptual Concepts. In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism*. Oxford University Press.
- Papineau, D. (2008). Explanatory Gaps and Dualist Intuitions. In L. Weiskrantz & M. Davies (Ed.), *Frontiers of Consciousness*. Oxford: Oxford University Press.
- Papineau, D. (2011). What Exactly is the Explanatory Gap? *Philosophia*, 39(1), 5-19.
- Pereboom, D. (2009). Consciousness and Introspective Inaccuracy. In L. Jorgensen & S. Newlands (Ed.), *Appearance, Reality, and the Good: Themes from the Philosophy of Robert M. Adams* (p. 156-187). Oxford University Press.
- Pereboom, D. (2011). *Consciousness and the Prospects of Physicalism*. Oxford University Press.
- Pohl, R. (2004). Introduction: Cognitive illusions. In R. Pohl (Ed.), *Cognitive Illusions. A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Hove, East Sussex: Psychology Press.
- Rey, G. (1995). Towards a Projectivist Account of Conscious Experience. In T. Metzinger (Ed.), *Conscious Experience*. Paderborn: Ferdinand Schöningh.
- Schroer, R. (2010). Where's the Beef? Phenomenal Concepts as Both Demonstrative and Substantial. *Australasian Journal of Philosophy*, 88(3), 505-522.
- Stoljar, D. (2005). Physicalism and phenomenal concepts. *Mind and Language*, 20(2), 296-302.
- Stoljar, D. (2006). *Ignorance and Imagination: The Epistemic Origin of the Problem of Consciousness*. Oxford: Oxford University Press.
- Sturgeon, S. (1994). The Epistemic View of Subjectivity. *The Journal of Philosophy*, 91(5), 221-235.
- Sturgeon, S. (2000). *Matters of Mind*. London: Routledge.
- Sundström, P. (2008). Is the mystery an illusion? Papineau on the problem of consciousness. *Synthese*, 163(2), 133-143.
- Sundström, P. (2011). Phenomenal Concepts. *Philosophy Compass*, 6(4), 267-281.

- Tversky, A., & Kahneman, D. (1983). Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review*, 90, 293-315.
- Tye, M. (1995). *Ten problems of consciousness : a representational theory of the phenomenal mind*. Cambridge (Mass.): the MIT Press.
- Tye, M. (1999). Phenomenal consciousness : the explanatory gap as a cognitive illusion. *Mind*, 108(432), 705-725.
- Wason, P. (1966). Reasoning. In B. Foss (Ed.), *New horizons in psychology 1*. Harmondsworth: Penguin.
- Wason, P., & Johnson-Laird, P. (1972). *Psychology of Reasoning : Structure and Content*. Cambridge (Mass.): Harvard University Press.